Assessing the Availability of the Preconception Component of High-Risk Obstetrical Services by Estimating the Use of Teratogenic Medications Before and During Pregnancy

Section 1. Basic Measure Information

1.A. Measure Name

Assessing the Availability of the Preconception Component of High-Risk Obstetrical Services by Estimating the Use of Teratogenic Medications Before and During Pregnancy

1.B. Measure Number

0233

1.C. Measure Description

Please provide a non-technical description of the measure that conveys what it measures to a broad audience.

Identifies the frequency with which teratogenic medications are dispensed to women before and during pregnancy.

1.D. Measure Owner

Collaboration for Advancing Pediatric Quality Measures (CAPQuaM).

1.E. National Quality Forum (NQF) ID (if applicable)

Not applicable.

1.F. Measure Hierarchy

Please note here if the measure is part of a measure hierarchy or is part of a measure group or composite measure. The following definitions are used by AHRQ':

1. **Please identify the name of the collection of measures to which the measure belongs (if applicable). A collection is the highest possible level of the measure hierarchy. A collection may contain one or more sets, subsets, composites, and/or individual measures.**

   This measure belongs to the Pediatric Quality Measures Program (PQMP) CAPQuaM's Availability of High-Risk Obstetric Services measure set.

2. **Please identify the name of the measure set to which the measure belongs (if applicable). A set is the second level of the hierarchy. A set may include one or more subsets, composites, and/or individual measures.**

   Availability of Preconception High-Risk Obstetric Care.

3. **Please identify the name of the subset to which the measure belongs (if applicable). A subset is the third level of the hierarchy. A subset may include one or more composites, and/or individual measures.**

   Teratogen subset.

4. **Please identify the name of the composite measure to which the measure belongs (if applicable). A composite is a measure with a score that is an aggregate of scores from other measures. A composite may include one or more other composites and/or individual measures. Composites may comprise component measures that can or cannot be used on their own.**

   Not applicable.

1.G. Numerator Statement

Various numerators are specified for the sub-measures in order to estimate the number of women who fill prescriptions for teratogenic medications in the specified circumstances, before and during pregnancy.

Numerator Elements:

Dated ICD-9 codes, DRGs, and prescription drug fill (or payment) data, including NDC codes (National Drug Codes) or compound names.

1.H. Numerator Exclusions

None.

1.I. Denominator Statement

Various denominators are specified for a series of sub-measures and include: overall number of deliveries; eligible qualifying high-risk pregnancies; and pregnancies with exposure to specified teratogenic (Class X) medications, using the indicated look-back period.

Eligible high-risk pregnancies look-back period, and specified teratogenic (Class X) medications are all described in detail in the measure specifications (see Supporting Documents).

Denominator Elements:

Maternal and infant ICD-9 codes.
Maternal DRG, CPT codes, and revenue codes.
Infant ICD-9 codes when available.
Pharmacy data, including NDC codes and/or compound names.

1.J. Denominator Exclusions

Denominator exclusions are identified using maternal ICD-9 codes specified in Table 1 (see Supporting Documents).

1.K. Data Sources

Check all the data sources for which the measure is specified and tested.

Administrative data (e.g., claims data).

If other, please list all other data sources in the field below.

Not applicable.

Section 2: Detailed Measure Specifications

Provide sufficient detail to describe how a measure would be calculated from the recommended data sources, uploading a separate document (+ Upload attachment) or a link to a URL. Examples of detailed measure specifications can be found in the CHIPRA Initial Core Set Technical Specifications Manual 2011 published by the Centers for Medicare & Medicaid Services. Although submission of formal programming code or algorithms that demonstrate how a measure would be calculated from a query of an appropriate electronic data source are not requested at this time, the availability of these resources may be a factor in determining whether a measure can be recommended for use.

See Supporting Documents for Table 1 and information on how to access the Technical Specifications for this measure.

Section 3. Importance of the Measure

In the following sections, provide brief descriptions of how the measure meets one or more of the following criteria for measure importance (general importance, importance to Medicaid and/or CHIP, complements or enhances an existing measure). Include references related to specific points made in your narrative (not a free-form listing of citations).

3.A. Evidence for General Importance of the Measure

Provide evidence for all applicable aspects of general importance:

- **Addresses a known or suspected quality gap and/or disparity in quality (e.g., addresses a socioeconomic disparity, a racial/ethnic disparity, a disparity for**

Children with Special Health Care Needs (CSHCN), a disparity for limited English proficient (LEP) populations).

- **Potential for quality improvement (i.e., there are effective approaches to reducing the quality gap or disparity in quality).**
- **Prevalence of condition among children under age 21 and/or among pregnant women.**
- **Severity of condition and burden of condition on children, family, and society (unrelated to cost).**
- **Fiscal burden of measure focus (e.g., clinical condition) on patients, families, public and private payers, or society more generally, currently and over the life span of the child.**
- **Association of measure topic with children's future health – for example, a measure addressing childhood obesity may have implications for the subsequent development of cardiovascular diseases.**
- **The extent to which the measure is applicable to changes across developmental stages (e.g., infancy, early childhood, middle childhood, adolescence, young adulthood).**

CAPQuaM was assigned the topic of availability of high-risk obstetrical services as a PQMP priority by AHRQ and the Centers for Medicare & Medicaid Services (CMS). Our measures were developed in close collaboration with our Expert Panel and our partner stakeholders.

Availability of specific aspects of care for pregnant women, particularly those in need of high-risk obstetric services, fosters healthy pregnancies and healthy deliveries. This measure describes availability of the preconception component of high-risk obstetrical care by estimating the use of teratogenic (i.e. "Class X") medications during potentially vulnerable periods before and during pregnancy as a marker for failures of availability. While preconception care for high-risk pregnancies is a broad topic, teratogen use in pregnancy offers a window that addresses a potential failure in the realm of patient safety.

This measure contains six sub-measures that describe how frequently a potentially dangerous circumstance occurs: pregnant women fill prescriptions for medications that have a very poor safety profile for pregnant women. The last (seventh) sub-measure describes the extent to which women who fill prescriptions for a teratogenic medication in the 15 months prior to pregnancy (from 2 years before delivery to the estimated time of conception) stop filling prescriptions for such medications during pregnancy.

This measure addresses an important topic (Kuhlthau, 2011). Medication use during pregnancy is common, with estimates ranging from < 30 percent to > 90 percent of women taking at least one prescription medication (Daw, Hanley, Greyson, et al., 2011; Mitchell Gilboa, Werler, et al., 2011). In the United States, women of reproductive age (15-44 years) receive nearly 12 million prescriptions for potentially teratogenic medications each year (Schwarz, Maselli, Norton, et al, 2005). A teratogen is a drug or substance capable of interfering with the development of a fetus

causing birth defects. The Food and Drug Administration (FDA) has a pregnancy category system (including categories A, B, C, D, and X) that describes the potential safety or risk of taking a medication during pregnancy (Table 2, see Supporting Documents). Due to limitations in the ability of the pregnancy categories to accurately and consistently convey the specific risk and benefit involved, the FDA has proposed new product labeling designed to improve risk versus benefit assessment of drugs used in pregnant and lactating women (FDA, 2014).

Both FDA Category D and X medications are considered potentially teratogenic because category D medications have evidence of human risk with benefits of the drug typically outweighing risks; category X medications have evidence of human risk with risk clearly outweighing the benefit. It is estimated that one woman in six (16 percent) of reproductive age fills a prescription for a class D or X medication (Schwarz, Postlethwaite, Hung, et al., 2007). Unfortunately, only 20-50 percent of these women receive contraceptive counseling at the time that medication is prescribed (Schwartz, et al., 2005, 2007). It has been estimated that one in four or 27 percent of U.S. pregnancies is exposed to potentially teratogenic medications that present potentially greater risk than benefit to the fetus (Schwarz, Parisi, Handler, et al., 2013; Schwarz, et al., 2007). Pregnant women use an average of 4.2 medications (OTC and prescription) throughout their pregnancy, with 93.9 percent taking at least one medication (Mitchell, et al., 2011).

Approximately half (51 percent) of the 6.6 million pregnancies each year in the United States are not planned (Guttmacher Institute, 2013). Thus, women may be exposed to potentially harmful teratogenic agents because they may not know they are pregnant or about to become pregnant; this is compounded because severe drug-induced malformations are more likely to occur within the first 3 months of pregnancy (Centers for Disease Control and Prevention [CDC], 2014; Guttmacher Institute, 2013). Assuring that women on high-risk drugs are off of those drugs before they get pregnant is a critical component of preconception care. Unplanned pregnancy rates are highest among young, poor, minority, and low-income women (Guttmacher Institute, 2013). Many women unknowingly expose their fetus to teratogens because they have not been counseled or managed sufficiently about teratogen use by their clinicians.

The most common teratogenic effects from medications include neural tube defects, congenital heart abnormalities, cleft lip or palate, and fetal stillbirth (Burkey, Holmes, 2013). Additional adverse fetal effects that result in dysfunction of a formed organ or tissue include postnatal adaptation, withdrawal, electrolyte abnormalities, and altered glucose metabolism (Lo, Friedman, 2002). Examples of currently used prescription medications with known risks of teratogenicity include angiotensin-converting enzyme inhibitors, carbamazepine, warfarin, methotrexate, phenytoin, isotretinoin, lithium, misoprostol, tetracycline, and valproate (Burkey, Holmes, 2013; Jentink, Dolk, Loane, et al., 2010; Jentink, Loane, Dolk, 2010).

Given the number and severity of fetal effects that can occur with many different medications, discussing medication use with women of reproductive age and child-bearing ability is critical. However, data suggest that only 20-50 percent of women receive contraceptive counseling when potentially teratogenic medications are prescribed (Gawron, Hammond, Keefer, 2014; Schwarz, et al., 2007; Schwarz, Longo, Zhao, et al., 2010; Schwarz, Santucci, Borrero, et al., 2009). In a survey of over 800 women, 43 percent of reproductive-age women prescribed potential teratogens reported no counseling from their provider about teratogenic risks (Schwarz, et al., 2013).

Another study demonstrated that among 146,758 women ages 18-44 years prescribed category X medications, only 26,136 (18 percent) also took oral contraceptives (Steinkellner, Chen Denison, 2010), a rate similar to that of same-aged women not taking category X medications (17 percent). The fact that many women do not receive appropriate counseling is concerning, as guidance from the Internet and other sources can be incorrect. In an environmental scan of 25 different Internet resources, including three medical and one professional organization, four pregnancy information resources, and 17 clinical practice resources, a total of 164 medication components were identified as "safe" for use by pregnant women (Peters, Lind, Humphrey, et al., 2013). When compared with the Teratogen Information System (TERIS), a database with expert assessments of the teratogenic risk of medication in human pregnancy after exposure, only 103 of those medications had existing evaluations, with 49 (48 percent) rated as unlikely to pose a risk and 43 (42 percent) of undetermined risk.

Prescription drug use during pregnancy is common. This measure considers the use of teratogenic medications during potentially vulnerable periods before and during pregnancy as a marker for insufficient availability of preconception and inter-conception high-risk obstetrical care.

3.B. Evidence for Importance of the Measure to Medicaid and/or CHIP

Comment on any specific features of this measure important to Medicaid and/or CHIP that are in addition to the evidence of importance described above, including the following:

- **The extent to which the measure is understood to be sensitive to changes in Medicaid or CHIP (e.g., policy changes, quality improvement strategies).**
- **Relevance to the Early and Periodic Screening, Diagnostic and Treatment benefit in Medicaid (EPSDT).**
- **Any other specific relevance to Medicaid/CHIP (please specify).**

The relevance of this measure for Medicaid and CHIP is demonstrated by a number of factors including: (1) the key Department of Health partnerships that played important roles in the development of the measure, (2) evidence demonstrating the high use of teratogen medications among pregnant women and the lack of appropriate counseling, and (3) the fact that unplanned pregnancies place women at higher risk for exposure to teratogens and that unplanned pregnancies are highest among young, minority, and low-income women who often are covered by Medicaid.

Our expert panel strongly endorsed the importance of preconception care and the use of exposure to teratogens as a marker of failure of adequate preconception care. Our consortium partners at the New York State Department of Health, including the Office of Health Insurance Programs / New York State Medicaid, steering committee, and scientific team also played central roles in the development of these measures. Evidence for a high level of interest in this work was demonstrated in particular by the fact that the CAPQuaM team was asked to present this work in development to the CMS Expert Panel on Improving Maternal and Infant Health Outcomes in Medicaid/CHIP Data, Measurement, and Reporting Workgroup. The New York State Office of Health Insurance Programs is an active CAPQuaM partner and has been engaged

in the conceptualization and development of these measures. Our testing occurred in Medicaid data and is described below.

As described in Section 3.A above, the literature confirms the importance of this measure for all women (Schwarz, et al., 2005, 2007; Taruscio, Arriola, Baldi, et al., 2014). Although the vast majority of studies on this topic have evaluated health maintenance organizations and survey databases to describe potentially teratogenic medications dispensed, contraceptive counseling, and pregnancy testing in this reproductive age group (Schwarz, et al., 2005, 2007; Taruscio, et al., 2014), a few studies have examined this topic among Medicaid enrollees. For example, one study evaluated a Medicaid program by analyzing category X prescriptions filled by 95,284 women enrolled in TennCare, Tennessee's program for Medicaid enrollees and individuals without health insurance. The researchers used administrative data and found that 391 women (4.1/1,000) filled a category X prescription during pregnancy (Cooper, Hickson, Ray, 2004). The most common medications filled were non-contraceptive estrogens (n=118 women; 1.24/1,000), sedatives (n=81 women; 0.85/1,000), and statins (n=71; 0.75/1,000), which represented 69 percent of all category X drug use. Most women (n=317; 81.1 percent) had a physician visit that was linked to the prescription. Furthermore, 239 (61.1 percent) of the 391 women filled a prescription for a category X drug > 28 days after the last menstrual period when pregnancy would have occurred, and 151 women (38.6 percent) filled a prescription for a category X drug after a physician visit in which pregnancy was diagnosed.

Certain subgroups of these women enrolled in TennCare had significantly increased risk of filling prescriptions for category X drugs during pregnancy. Subgroups included those above the age of 35 years and those enrolled in TennCare because of disability. Those older than 35 years of age had nearly 10 times increased risk of filling a prescription for category X drugs during pregnancy as compared to women younger than 18 years of age, 12.1 versus 1.5 per 1,000, respectively ($p<0.0001$). And those enrolled in TennCare because of disability were nearly three times as likely to fill prescriptions for category X medications during pregnancy as women in other categories ($p<0.0001$).

This study did not include reproductive age women of child-bearing ability that were not pregnant and exposed to potentially teratogenic medications or pregnant women exposed to category D medications. Therefore, highly vulnerable groups and additional fetal risks were not evaluated. The authors emphasized the need for monitoring of care delivery and communication to providers regarding exposure of women to category X medications. Specifically, communication should be focused on the highest risk populations, including older reproductive age women and those with chronic health conditions.

Another study evaluated a small subset of women (n=105) enrolled in Michigan Medicaid who took loperamide in pregnancy, a drug with unknown effects at that time. It was found that the use of loperamide during pregnancy was not associated with an increased risk of major malformations. In a very recent study, a large national Medicaid database was used to examine rates of cardiac malformations with first-trimester antidepressant exposure in nearly 950,000 pregnant women (Huybrechts, Palmsten, Hernandez-Diaz, et al, 2014). Cardiac malformations were diagnosed in 90 per 10,000 antidepressant-exposed infants versus 72 per 10,000 unexposed infants, resulting in no significant difference between groups. While these studies had negative

findings, they support the use of Medicaid databases for data analysis and the importance of evaluating this population for potential teratogenic risk.

Overall, although information evaluating pregnant women taking potentially teratogenic medications enrolled in Medicaid or Children's Health Insurance Program (CHIP) programs is inadequate, findings above confirm the importance of this measure for all women. There is widespread use of teratogens by women of reproductive age, and half of all pregnancies are not planned making this measure of high relevance. These issues are particularly acute for low-income women of color, a population disproportionately covered by Medicaid. High-risk pregnancies are common in general and more so in Medicaid populations.

3.C. Relationship to Other Measures (if any)

Describe, if known, how this measure complements or improves on an existing measure in this topic area for the child or adult population, or if it is intended to fill a specific gap in an existing measure category or topic. For example, the proposed measure may enhance an existing measure in the initial core set, it may lower the age range for an existing adult-focused measure, or it may fill a gap in measurement (e.g., for asthma care quality, inpatient care measures).

Previously, we developed measures based on institutional self-report of whether there is 24 hour 7 day a week availability of structural characteristics at the facility in which the woman gave birth. We also developed two measures that focused on the availability of specialty physician services and multidisciplinary care for high-risk pregnant women. The current measure focuses on teratogenic drug exposure as a marker of failure of availability of services. It will supplement the collection of measures focused on HROB services to further evaluate and enhance the safety and care for high-risk women regardless of birth outcome. This measure represents a measure of safety for mother and infant. As appropriate, definitions and diagnoses used for this measure are harmonious with those of other CAPQuaM HROB measures.

The selection of these topics is valid and justified by evidence summarized above. All were prioritized during our formal expert process.

Key Recommendations from International and National Projects

The importance of this measure is highlighted by two recent reports from the EUROCAT and EUROPLAN projects and the CDC (Broussard, Frey, Hernandez-Diaz, et al., 2014; Taruscio, et al., 2014) The two European projects have joined together to provide policy recommendations for primary prevention of congenital anomalies. The recommendations include interdisciplinary expertise to encompass different actions aimed at reducing risk factors and increasing protective factors and behaviors. The scope of actions includes the field of medicinal drugs and specifically outlines goals of:

- Advising women to seek medical advice before trying to get pregnant.
- Ensuring guidelines will be made available for physicians regarding the risk-benefit balance for use of medications in pregnancy (particularly those used for treating chronic diseases).
- Providing a teratogen information service where specialized advice can be obtained by

women and professionals; and conducting post-marketing pharmaco-vigilence to detect any risk of congenital anomalies associated with the use of medications

In addition, the CDC recently solicited expert input on an outline for a systematic approach to evaluating the quality and strength of evidence for associated risks of medication use in pregnancy. This strategy is known as “Treating for Two: Safer Medication Use in Pregnancy Initiative” (CDC, 2014). It aims to identify birth defects prevention and optimize maternal health by improving clinical decisions about management of common conditions in pregnancy and in the reproductive years. The proposed review will incorporate an evidence synthesis and review as well as guideline development via an independent panel of clinical, public health, and prevention experts. Primary outcomes that would be evaluated include preterm birth, fetal death, structural birth defects, poor fetal growth, neurocognitive and behavioral effects, and severe adverse maternal events. This multidisciplinary panel of experts proposed that this prioritization, synthesis, evaluation, and dissemination of safety information is of high clinical and public health relevance.

Our measure complements this focus. We suggest availability of care is essential for women who are at risk of adverse pregnancy outcomes due to teratogenic exposure. This measure has the potential to improve perinatal outcomes in the setting of high-risk pregnancies. Thus, this measure strives to decrease the number of pregnancies exposed to teratogenic drugs. Further, our measure also assesses a critical component of safety for this population, as high-risk women with inadequate preconception and inter-conception care represent a critical failure of the system.

Section 4. Measure Categories

CHIPRA legislation requires that measures in the initial and improved core set, taken together, cover all settings, services, and topics of health care relevant to children. Moreover, the legislation requires the core set to address the needs of children across all ages, including services to promote healthy birth. Regardless of the eventual use of the measure, we are interested in knowing all settings, services, measure topics, and populations that this measure addresses. These categories are not exclusive of one another, so please indicate "Yes" to all that apply.

Does the measure address this category?

a. **Care Setting – ambulatory:** Yes.
b. **Care Setting – inpatient:** No.
c. **Care Setting – other – please specify:** No.
d. **Service – preventive health, including services to promote healthy birth:** Yes.
e. **Service – care for acute conditions:** Yes.
f. **Service – care for children with special health care needs/chronic conditions:** No.
g. **Service – other (please specify):** No.
h. **Measure Topic – duration of enrollment:** No.
i. **Measure Topic – clinical quality:** Yes.
j. **Measure Topic – patient safety:** Yes.
k. **Measure Topic – family experience with care:** No.

l. **Measure Topic – care in the most integrated setting:** No.
m. **Measure Topic other (please specify):** No.
n. **Population – pregnant women:** Yes; ages 10 to 65 years.
o. **Population – neonates (28 days after birth) (specify age range):** No.
p. **Population – infants (29 days to 1 year) (specify age range):** No.
q. **Population – pre-school age children (1 year through 5 years) (specify age range):** No.
r. **Population – school-aged children (6 years through 10 years) (specify age range):** Yes; pregnant >=10 years of age.
s. **Population – adolescents (11 years through 20 years) (specify age range):** Yes; pregnant, ages 11-20 years.
t. **Population – other (specify age range):** Yes; pregnant <=65 years of age.
u. **Other category (please specify):** No.

Section 5. Evidence or Other Justification for the Focus of the Measure

The evidence base for the focus of the measures will be made explicit and transparent as part of the public release of CHIPRA deliberations; thus, it is critical for submitters to specify the scientific evidence or other basis for the focus of the measure in the following sections.

5.A. Research Evidence

Research evidence should include a brief description of the evidence base for valid relationship(s) among the structure, process, and/or outcome of health care that is the focus of the measure. For example, evidence exists for the relationship between immunizing a child or adolescent (process of care) and improved outcomes for the child and the public. If sufficient evidence existed for the use of immunization registries in practice or at the State level and the provision of immunizations to children and adolescents, such evidence would support the focus of a measure on immunization registries (a structural measure).

Describe the nature of the evidence, including study design, and provide relevant citations for statements made. Evidence may include rigorous systematic reviews of research literature and high-quality research studies.

Every 4.5 minutes a baby is born with a birth defect in the United States: 120,000 babies are affected by birth defects each year (CDC, 2014). Certain medications increase the risk of having a birth defect. Negative outcomes can result from fetal exposure to potentially teratogenic medications. This risk can be minimized or avoided with reduced exposure to these potentially teratogenic medications. The use of highly effective contraception (i.e., intrauterine devices, implant, sterilization) can avoid the unplanned concurrence of pregnancy with teratogen use. The most effective structures and processes to provide information to clinicians and to patients about which medications have potential risk and how to communicate risk minimization strategies have not as yet been fully elucidated.

Teratogenic risk counseling is not universal, with the majority of patients reporting no counseling by their clinician despite being prescribed teratogenic medications. For example, only 40 percent of individuals prescribed carbamazepine and 22 percent prescribed valproate receiving this counseling (Langan, Perry, Oto, 2013). Only 13-17 percent of women had documentation surrounding contraceptive issues. In general practice, less than 60 percent of physicians correctly identify category D or X medications, and only 65 percent understand contraceptive failure rates for various contraceptive methods (Eisenberg, Stika, Desai, 2010). This is especially concerning when 1 in every 13 ambulatory care visits involves a prescription for a teratogenic medication (Schwarz, et al., 2005). Physicians agree they should be providing information about contraception and teratogenic medications, and patients want that information, so it is critical to ensure that appropriate information and appropriate contraceptive counseling methods are in place (Eisenberg, et al., 2010; Santucci, Gold, Akers, et al., 2010; Schwarz, et al., 2009). Counseling to avoid the concurrence of pregnancy and teratogen use is a fundamental aspect of preconception care.

While clinicians do feel responsible for counseling women about risks when they prescribe medications that may cause birth defects, they perceive many barriers that prevent them from doing so (Akers, Gold, Borrero, et al., 2010; Schwarz, et a., 2009). Changes in alerts within electronic health records or clinical decision support systems have not been effective (Adam, Polifka, Friedman, 2011; Schwarz, et al., 2013; Schwarz, Parisi, Handler, et al., 2012). The transition within the FDA from pregnancy categories to labeling requirements, which will describe in more detail human effects in pregnancy (and lactation) for individual medications, may cause additional confusion and provide information that is difficult for clinicians to interpret. While documentation of contraception as a vital sign and quality improvement interventions in primary care have been suggested to address using contraception when potentially teratogenic medications are prescribed, there has been virtually no adoption of effective and sustainable methods that minimize risk of teratogenicity in reproductive age women taking teratogenic medications (Force, Keppel, Guirguis-Blake, et al., 2012; Schwarz, et al., 2010). This emphasizes a critical need to identify and develop effective and sustainable ways to reduce teratogenic risk.

For every reproductive age woman, the benefits and risks of medications must be weighed. For women with epilepsy or depression, the benefit of remaining on a potentially teratogenic medication may outweigh the medication risk. Preconception planning and counseling may minimize risk to the fetus and the mother. In other circumstances when pregnancy is not desired and these potentially teratogenic medications are used, providing a highly effective form of contraception may avoid teratogenic risk. Thus, while extenuating circumstances may favor use of teratogenic medications during pregnancy, this should be rare.

This measure represents a marker for the availability of the preconception component of high-risk obstetrical care as it estimates the use of teratogenic medications during potentially vulnerable periods before and during pregnancy as a marker for failures of availability. This measure is of high relevance given the frequency of teratogenic medication use among pregnant women before and during pregnancy.

5.B. Clinical or Other Rationale Supporting the Focus of the Measure (optional)

Provide documentation of the clinical or other rationale for the focus of this measure, including citations as appropriate and available.

This is discussed in detail in sections above. Birth defects and other adverse fetal effects due to medication use are preventable with proper identification and action. The ability to minimize and/or prevent adverse fetal effects due to potentially teratogenic medication use support the focus of this measure with identification of at-risk populations. Given that approximately half of pregnancies are unplanned, this measure will promote patient safety to both the mother and fetus and reduce the potential for adverse outcomes. Appropriate availability of specific aspects of care for women is necessary to achieve desired pregnancy outcomes, including delaying the onset of pregnancy until a time when it is safe for the woman to become pregnant. This measure describes availability of the preconception component of high-risk obstetrical care by identifying the use of teratogenic medications during potentially vulnerable periods before and during pregnancy as evidence of a specific failure of preconception care. Since birth defects have financial as well as health costs, this measure relates to all six characteristics (Timely, Equitable, Safe, Efficient, Patient-Centered and Effective) of quality care described in the Institute of Medicine's report, Crossing the Quality Chasm (2001). We have described the importance of this measure in our review above. The proposed measure can provide a marker for the availability of preconception care.

The salience and validity of our work has benefited from our use of a formal method, a pragmatic adaptation of the CAPQuaM 360 degree method. The method, as adapted to availability of HROB services, described in the next paragraph was specifically designed to develop valid and reliable measures in the face of pragmatic epistemological uncertainty. That is, recognizing that practice extends well beyond the research base, we designed this method to allow us to develop reliable and valid state of the science measures, in part by explicitly modeling and accounting for uncertainties in the measure development, in part by the conceptualization and implementation of a Boundary Guideline. We have shared and refined this approach in a number of venues, including within the PQMP, which comprises the various PQMP AHRQ-CMS CHIPRA Centers of Excellence, the State PQMP participants, and AHRQ and CMS participants. All presentations have invited dialogue and feedback. This work has been similarly presented at a number of Grand Rounds / weekly conferences in the New York-New Jersey area as well as to national/international audiences, including the bioethics and children's health services communities. These latter venues include:

- 2012 Pediatric Academic Societies State of the Science Plenary (Boston).
- 2012 Oxford-Mount Sinai Bioethics Consortium (Amsterdam).
- 2012 Child Health Services Research Interest Group at Academy Health (Orlando).

Feedback from these presentations has been extremely positive. The Boundary Guideline construct has generated particular enthusiasm. We asked the Bioethics Consortium to extrapolate the primum non nocere (First, do no harm) principle to apply regarding this aspect of performance measurement. We received strong feedback that not only is it ethical to measure using

systematically developed measures (even in the context of some uncertainty), but that it is ethically preferable to use such measures compared with the alternative of providing care that is not assessed (and perhaps not assessable) because of residual uncertainty. Fortunately, in the case of this proposed measure, we can present both a systematically developed measure and a variety of evidence to support its use.

Section 6. Scientific Soundness of the Measure

Explain the methods used to determine the scientific soundness of the measure itself. Include results of all tests of validity and reliability, including description(s) of the study sample(s) and methods used to arrive at the results. Note how characteristics of other data systems, data sources, or eligible populations may affect reliability and validity.

6.A. Reliability

Reliability of the measure is the extent to which the measure results are reproducible when conditions remain the same. The method for establishing the reliability of a measure will depend on the type of measure, data source, and other factors.

Explain your rationale for selecting the methods you have chosen, show how you used the methods chosen, and provide information on the results (e.g., the Kappa statistic). Provide appropriate citations to justify methods.

The strengths of this measure derive from its systematic development, its thoughtful specification, its careful conceptualization and articulation, and its grounding in existing science and consensus.

Our specifications rely on the use of administrative data. These data are used to identify deliveries (our specifications are based on CDC methodologies described in Kuklina and Callaghan, 2011). We initially tested these specifications in Medicaid MAX data. We added Revenue Code 722 to Kuklina's specified list, as the Medicaid MAX data provided by CMS do not include DRGs, which are employed in the Kuklina method. We also tested a variation of the approach to identify deliveries employed by HEDIS in its Timing of Prenatal Care measure in the initial CHIPRA core set. We found that these approaches identified substantially the same population of deliveries in a 16-State subset of the national MAX database. We chose the16 States to include in an attempt to manifest some standardization of approaches across the seven AHRQ-CMS CHIPRA Centers of Excellence—they were recommended to us as a diverse set of States with high data quality by the Children's Hospital of Pennsylvania Center (CHOP), which has used them extensively in a number of their validation activities. As the different approaches produced 90 percent or more overlap, we decided to specify the measure based on the Kuklina/CDC approach as both widely used and relevant for the type of population-based approach to measurement proposed in this measure. We have used this method for all of the CAPQuaM high-risk obstetrical services availability measures.

In determining which women were to be considered potentially in need of HROB services, our specifications further rely on administrative data. One study found that quality measures that could be calculated using administrative data showed higher rates of performance than indicated

by a review of the medical record alone, and that claims data are more accurate for identifying services with a high likelihood of documentation due to reimbursement (Diamond, Rask, Kohler, 2004). Further, at the current stage of electronic health record (EHR) development and implementation, chart review is likely to prove infeasible for population-based measures of this scope. Since we identify all women for whom the prescription data indicate teratogen use during the year before delivery, we are unlikely to miss very many women who were provided prescription teratogenic medications within their insurance plan. We found that of ~119,000 Medicaid deliveries in New York State in 2010, 59,254 were at sufficiently elevated risk to qualify for this measure set (approximately 50 percent). Our team had predicted that 50 to 60 percent of all pregnancies would have elevated risk. Use of a mother-only algorithm in MAX data in 16 States indicates the proportion of high-risk pregnancies ranges from 31.50 percent in New Jersey to 63.97 percent in Kentucky. The New York MAX finding was 55,379 HROB pregnancies, almost identical to the 56,465 found using internal data bases on the maternal codes, indicating very high reliability across systems. In the New York State 2012 mother-baby linked dataset used for testing the final specifications, of 102,399 linked files 61,676 (60 percent) qualified as HROB. The vast majority of the sample (82 percent) qualified because of maternal comorbid illness or because of pregnancy complications. This is consistent with our estimate that 50 to 60 percent of Medicaid pregnancies would be at elevated risk and within the range seen in our previous 16-State analysis.

As for the specification of teratogens, our scientific team, including an expert pharmacist, generated a list of class X teratogens for the testing of this measure. Databases searched for classification included the FDA, Micromedix, and Reprotox databases. Our list was conservative including only Class X medications and omitting things such as Hydroxyzine HCL, an antihistamine suspected of causing birth defects and for which safer in class medications are available. Other medications might prove fertile for future testing for measurement. In our testing of the proportion of women who fill prescriptions for Class X medications before and during pregnancy, 323 deliveries (0.32 percent) filled prescriptions within the 9 months prior to delivery, 1,167 deliveries (1.14 percent) filled prescriptions within the 12 months prior to delivery, and 3,405 deliveries (3.33 percent) filled prescriptions within the 24 months prior to delivery.

In contrast to our hope that this would be a rare event, data analyses completed in 2012 New York State Medicaid linked mother-baby records suggest that teratogen use within 12 months of pregnancy occurs with some frequency, more so among pregnancies identified using our HROB algorithm. Although a significant percentage of women who filled two or more prescriptions within 24 months of delivery stopped using teratogens by 9 months prior to their delivery, there was still a considerable number who filled prescriptions for a teratogenic medication during pregnancy; Table 3 (see Supporting Documents) describes results from testing this measure in New York State 2012 Medicaid data.

6.B. Validity

Validity of the measure is the extent to which the measure meaningfully represents the concept being evaluated. The method for establishing the validity of a measure will depend on the type of measure, data source, and other factors.

Explain your rationale for selecting the methods you have chosen, show how you used the methods chosen, and provide information on the results (e.g., R2 for concurrent validity).

The reliability section above also contains information related to validity.

Our definition of high-risk obstetrical services results from a formal RAND/UCLA modified Delphi process conducted with a multidisciplinary panel of national experts that included obstetricians, maternal-fetal medicine (MFM) specialists, a nurse midwife, an anesthesiologist, and a family physician. We carefully operationalized the panel's clinical recommendations by fine tuning AHRQ's Clinical Classification Software. We operationalized panel specifications using data elements that are available in typical administrative data sets.

Potential exceptions are elements such as race and ethnicity. Our feasibility work confirmed race/ethnicity data are generally available in Medicaid datasets. The CHIPRA legislation (2009) directs our measures to be capable of identifying disparities, and we have specified it to be so, although we are aware of variability in the manner of assignment of race and ethnicity by health care facilities.

Use of administrative data in performance assessment is common. Such data contain consistent elements, are available, inform regarding large numbers of individuals, and are relatively inexpensive. The validity of many administrative datasets has been established, and their strengths and weaknesses relative to data abstracted from medical records and obtained via survey have been documented. Their use has been encouraged by Federal agencies (Virnig, McBean, 2001). The Centers for Medicare & Medicaid Services (CMS) has made clear to the participating AHRQ-CMS CHIPRA Centers of Excellence funded to develop measures in the Pediatric Quality Measures Program that it places a premium on feasibility.

The ability of expert panels to enhance measure development and health care evaluation, including for children, has been demonstrated (Mangione-Smith, DeCristofaro, Setodji, et al., 2007). Frontline practitioners can assist researchers in creating useful measures (Rubio, Berg-Weger, Tebb, et al., 2003). CAPQuaM's 360 degree method is highly engaged with collaborators, partners, and the literature. It targets relevant information and perspective, and measures emerge from the process. Potential measures are tested to the extent that time and resources permit. In developing the HROB availability measures we incorporate:

- Engagement with broadly diverse partnered institutions and senior advisors.
- Detailed literature review.
- Interviews with clinicians from around the country.
- The CAPQuaM scientific team.
- A geographically diverse, multidisciplinary expert panel that participated in a two round RAND/UCLA modified Delphi process, with enhanced follow-up.
- Development of a Boundary Guideline that incorporates simultaneously a variety of gradients, including gradients of importance, relevance, and certainty, as appropriate to the construct being represented.

- Specification and review of measures and approaches to measurement by stakeholders and experts.
- Testing and assessment of measure performance using Medicaid data.

Availability

The construct of availability is complex and can be muddied in the distinction or lack thereof between availability, access, and utilization. For this PQMP measure set on availability of preconception HROB services, we use teratogen use as a marker of failure of availability of preconception (inter-conception) care among women needing HROB services. While these measures are challenging to validate definitively, evidence of systematic variation may suggest construct validity.

High Risk

We have operationalized a systematic expert process informed by a detailed literature review and incorporating a well-described and frequently used system developed by AHRQ. While we have modified this system, it has been done to be consistent with its use in this context and to remain consistent with the guidance of the expert panel. It is transparent and has high face validity. We validated its use in 16 States using MAX data and in two separate years of New York State Medicaid data. The rate of HROB ranges from 50 to 60 percent across these datasets, the results are consistent with the fact that Medicaid pregnancies are higher risk because of higher rates of comorbid illness and pregnancy complications as demonstrated in the literature.

Teratogen Use as a Marker for Availability of Preconception Care

This measures describes availability of the preconception component of high-risk obstetrical care by using the use of teratogenic (i.e. Class X) medications during potentially vulnerable periods before and during pregnancy as a marker for failures of availability. While only a small slice of what may be accomplished with preconception care, avoiding the use of teratogenic medications during pregnancy (or delaying pregnancy while such medications are in use) has face validity, was advocated by our expert panel, and may be a population marker for insufficient access to high quality preconception care for women at increased risk.

This construct was endorsed by our expert panel, and our list of teratogenic medications was generated by our scientific team, including obstetricians and pharmacists. Our list of teratogenic medications includes FDA pregnancy category system (Class X) medications, which are considered potentially teratogenic because they have evidence of human risk clearly outweighing the benefit. Pregnancy category classification was generated from the FDA, Micromedix, and Reprotox databases. Although Class D medications can be considered teratogenic, our expert panel did not endorse the inclusion of these medications in this measure as Class D medications are more likely to be used appropriately in high-risk pregnancies.

Section 7. Identification of Disparities

CHIPRA requires that quality measures be able to identify disparities by race, ethnicity, socioeconomic status, and special health care needs. Thus, we strongly encourage

nominators to have tested measures in diverse populations. Such testing provides evidence for assessing measure's performance for disparities identification. In the sections below, describe the results of efforts to demonstrate the capacity of this measure to produce results that can be stratified by the characteristics noted and retain the scientific soundness (reliability and validity) within and across the relevant subgroups.

7.A. Race/Ethnicity

Data elements on race and ethnicity are often available in administrative data and are typically available to Medicaid programs. The New York State Medicaid Program was able to identify race using their information systems; 45 individuals out of nearly 60,000 pregnancies were missing data on race.

We examined race/ethnicity data in New York State Medicaid files for 2012. Of the over 102,399 deliveries included in our testing, 12,803 (21 percent) were black, non-Hispanic; 18,459 (30 percent) were Hispanic; 18,965 (31 percent) were white; and 11,486 (195) were other. We found that approximately 1 percent of all deliveries, and 1.8 percent of HROB deliveries, filled prescriptions for teratogenic (Class X drugs) during the 12 months prior to their delivery date. Table 4 (see Supporting Documents) describes the proportion of women who filled prescriptions for Class X medications within the 12 months prior to their delivery date by race.

Pairwise comparisons revealed significant differences by race for teratogen use within 12 months of delivery (whites versus all other races, p=.02; blacks versus all other races, NS; Hispanics versus all other races, p<.001; others versus all other races, p=.001). These results suggest that variation in teratogen use can be identified with this measure. We recommend that reporting of results using this measure be stratified by race and ethnicity.

7.B. Special Health Care Needs

As a group, women with high-risk pregnancy constitute a population with special health care needs, although they are not strictly "children with special health care needs."

7.C. Socioeconomic Status

We used the national distribution of percent of individuals in poverty to establish five categories that reflect the degree of poverty at the county level. We considered other data, such as county median income or county unemployment, but felt that the percent of individuals in poverty was a more integrative measure. The use of a geographic rather than an individual measure is consistent with recent applications of hierarchical methods to study the impact of poverty and also with data that indicate that local disparities in income are an independent predictor of outcomes. It also allows this measure to consider issues of socioeconomic status while using publicly available data and requiring only the mother's county of residence, a more reliable data point than self- reported income.

Our analysis of U.S. Department of Agriculture (USDA) data considering 3,142 counties and related geographic units found a mean of 17.2 percent of county residents living in poverty, a standard deviation of 6.5 percent, and an interquartile range of 8.2 percent. The distribution illustrated below, shows meaningful dispersion and supports our plan to build off quartiles of

distribution with a finer focus in higher areas of poverty (see Table 5 in the Supporting Documents).

All of New York State lies in the top three quartiles. We would expect to find the largest differences between poorer and other counties than across the upper end of the spectrum. Plans can use county poverty levels to stratify measures by level of poverty.

7.D. Rurality/Urbanicity

As described in the specification, we used the Urban Influence Codes (UIC) below to describe the level of rurality or urbanicity.

Metropolitan

1. In large metro area of 1+ million residents
2. In small metro area of less than 1 million residents

Non-metropolitan

3. Micropolitan adjacent to large metro
4. Non-core adjacent to large metro
5. Micropolitan adjacent to small metro
6. Non-core adjacent to small metro with own town
7. Non-core adjacent to small metro no own town
8. Micropolitan not adjacent to a metro area
9. Non-core adjacent to micro with own town
10. Non-core adjacent to micro with no own town
11. Non-core not adjacent to metro or micro with own town
12. Non-core not adjacent to metro or micro with no own town

We analyzed 3,143 U.S. county equivalents; the results are presented in Table 6 (see Supporting Documents). The population is heavily weighted to metropolitan areas, as shown in Table 7 (see Supporting Documents).

As noted, we use Urban Influence Codes (UIC), which have been developed by the USDA based on a number of criteria to describe the levels of urbanicity and rurality. This is intended not only to report within plan differences but also to allow for aggregation as appropriate. While each UIC has its own meaningful definition, some researchers choose to aggregate various codes. We recommend consideration of the aggregation schema of Bennett and colleagues at the South Carolina Rural Research Center (Bennett, Olatosi, Probst, 2008). Their aggregation scheme brings together Codes 1 and 2 as Urban; 3, 5, and 8 as micropolitan rural; 4, 6, and 7 as rural adjacent to a metro area; and 9, 10, 11, and 12 as remote rural. We observe that UIC 5 might also be aggregated with 4, 6, and 7 as an adjacent rural area. Further, this approach to rurality does not map exactly to the population density based definition of frontier (fewer than six persons per square mile) as articulated in the Affordable Care Act (ACA). However, use of such categories is

consistent with the ACA's intent that the Secretary ask that data collected for racial and ethnic disparities also look at underserved frontier counties. Frontier health care may be approximated by analysis of the remote rural categories (Hart, 2012). Our judgment was confirmed after CAPQuaM consulted with Gary Hart, Director of the Center for Rural Health at the University of North Dakota School of Medicine & Health Sciences, who led a Health Resources and Services Administration (HRSA)-funded project to develop new methods to analyze frontier health. We clarified that his work suggests that UIC 9-12 is the best overall approach to using county level data to study frontier health. Those interested in care specific to large cities may wish to aggregate rural areas and analyze UIC 1 and 2 separately. Our testing suggests confirms that analyzing UIC 1 and 2 separately is necessary for certain settings, and the findings considering overall teratogen use suggest this.

The New York State Medicaid data were less sensitive to the distinction between urban versus rural; but the data were sensitive to the distinction between large versus smaller urban cities. Table 8 (see Supporting Documents) describes the proportion of women who fill prescriptions for Class X medications within 12 months prior to their delivery date by Urban Influence Code.

Pairwise comparisons revealed that UIC 1 had higher rates of teratogen use as compared with UIC 2-9 ($p<.0001$), and UIC 2 had lower rates of teratogen use as compared with UIC 1, 3-9 ($p<.0001$). These results may suggest that practice patterns vary between large urban cities and smaller cities.

7.E. Limited English Proficiency (LEP) Populations

Not assessed, but there is nothing intrinsic to the measure to inhibit its use in that population so long as the LEP characteristic can be linked to the pregnancy or delivery data.

Section 8. Feasibility

Feasibility is the extent to which the data required for the measure are readily available, retrievable without undue burden, and can be implemented for performance measurement. Using the following sections, explain the methods used to determine the feasibility of implementing the measure.

8.A. Data Availability

1. What is the availability of data in existing data systems? How readily are the data available?

The definitions were specified to allow their use with data elements that are available in administrative data, including health plan or State Medicaid programs. While zip code is sometimes a hidden or non-public variable when such data sets are released, it generally is available to a responsible entity, such as an insurer or a Medicaid program. Race and ethnicity are typically available to Medicaid programs and are on institutional medical records (e.g., hospital records). They are often but not always recorded in insurance databases. We have data from a feasibility study conducted confirming that both data elements are generally available in

the medical record, frequently electronically. The rapid expansion of data gathering from electronic health records can help augment administrative data review in measure assessment.

The CAPQuaM High-Risk OB measures seek to assess the proportion of high-risk women that are exposed to teratogens (Class X medications) during their pregnancies. As such, the data elements of interest include:

- Prescription fill date.
- Outpatient claims data.
- Documentation of conditions that would classify a woman as "high risk" for stratification purposes.
- Race and ethnicity.
- Insurance type (Medicaid, private, uninsured).
- Managed care insurance—Yes, No (where applicable).
- Benefit category (for Medicaid- and CHIP-eligible cohorts).
- Income level (as recorded for Medicaid- and CHIP-eligible cohorts).
- County equivalent and State or zip code of residence.

Several of these data elements are readily available through hospital administrative data. For example, identification of women with "high risk" conditions can be achieved through use of the appropriate ICD-9, CCS, and/or revenue codes. Prescription fill data can be achieved through administrative data. Additionally, benefit type is typically recorded in health plan, Medicaid, and CHIP administrative data sets.

As part of our feasibility assessment, CAPQuaM partnered with New York State Medicaid to conduct a variety of analyses using their administrative dataset. The findings from these analyses indicate that the administrative data elements are readily available at the State level and can be abstracted and used for calculating and reporting the CAPQuaM HROB measures. Further, we have specified several variables, including socioeconomic status (SES) and urbanicity, by linking county of residence at the time of delivery to publicly available data sets.

Payment source (insurance type) should be available in a health plan database and is also easily obtained from administrative data.

2. If data are not available in existing data systems or would be better collected from future data systems, what is the potential for modifying current data systems or creating new data systems to enhance the feasibility of the measure and facilitate implementation?

The data required for the CAPQuaM HROB measures are generally available in the existing data systems. Enhancement of linkages between mothers and babies, the routine reporting of estimated gestational age at delivery or date of conception, and similar data infrastructure would extend the capacity for refined reporting of these sort of HROB measures. Enhancement of collection of

patient-reported race and ethnicity data into existing administrative systems would also be valuable.

8.B. Lessons from Use of the Measure

1. Describe the extent to which the measure has been used or is in use, including the types of settings in which it has been used, and purposes for which it has been used.

New measure.

2. If the measure has been used or is in use, what methods, if any, have already been used to collect data for this measure?

The measure is not currently in use.

3. What lessons are available from the current or prior use of the measure?

The measure is not currently in use.

Section 9. Levels of Aggregation

CHIPRA states that data used in quality measures must be collected and reported in a standard format that permits comparison (at minimum) at State, health plan, and provider levels. Use the following table to provide information about this measure's use for reporting at the levels of aggregation in the table.

For the purpose of this section, please refer to the definitions for provider, practice site, medical group, and network in the Glossary of Terms.

If there is no information about whether the measure could be meaningfully reported at a specific level of aggregation, please write "Not available" in the text field before progressing to the next section.

Level of aggregation (Unit) for reporting on the quality of care for children covered by Medicaid/ CHIP†:

State level* Can compare States

Intended use: **Is measure intended to support meaningful comparisons at this level? (Yes/No)**
Yes.

Data Sources: **Are data sources available to support reporting at this level?**
Yes.

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**

Minimum size specified for analysis is 250. A study of HROB deliveries in MAX data in 18 States using slightly less sensitive criteria than those specified herein found they range from 1,637 (Vermont) to 55,382 (New York). The median is 14,500, with 25 percent less than 4,000 deliveries.

In Use: **Have measure results been reported at this level previously?**
No.

Reliability & Validity: **Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?**
No.

Unintended consequences: **What are the potential unintended consequences of reporting at this level of aggregation?**
None anticipated.

Other geographic level: Can compare other geographic regions (e.g., MSA, HRR)

Intended use: **Is measure intended to support meaningful comparisons at this level? (Yes/No)**
Yes.

Data Sources: **Are data sources available to support reporting at this level?**
Yes.

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**
Minimum size specified for analysis is 250. Study of HROB deliveries in MAX data in 18 States using slightly less sensitive criteria than those specified herein found they range from 1,637 (Vermont) to 55,382 (New York). The median is 14,500, with 25 percent less than 4,000 deliveries. We specify using urban influence codes, which allows for a variety of analyses.

In Use: **Have measure results been reported at this level previously?**
No.

Reliability & Validity: **Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?**
No.

Unintended consequences: **What are the potential unintended consequences of reporting at this level of aggregation?**
None anticipated.

Medicaid or CHIP Payment model: Can compare payment models (e.g., managed care, primary care case management, FFS, and other models)

Intended use: **Is measure intended to support meaningful comparisons at this level? (Yes/No)**
Yes.

Data Sources: **Are data sources available to support reporting at this level?**
Yes.

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**
Minimum size specified for analysis is 250. Study of HROB deliveries in MAX data in 18 States using slightly less sensitive criteria than those specified herein found they range from 1,637 (Vermont) to 55,382 (New York). The median is 14,500, with 25 percent less than 4,000 deliveries.

In Use: **Have measure results been reported at this level previously?**
No.

Reliability & Validity: **Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?**
No.

Unintended consequences: **What are the potential unintended consequences of reporting at this level of aggregation?**
None anticipated.

Health plan*: Can compare quality of care among health plans.

Intended use: **Is measure intended to support meaningful comparisons at this level? (Yes/No)**
Yes.

Data Sources: **Are data sources available to support reporting at this level?**
Yes.

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**
Minimum size specified for analysis is 250. Study of HROB deliveries in MAX data in 18 States using slightly less sensitive criteria than those specified herein found they range from 1,637 (Vermont) to 55,382 (New York). The median is 14,500, with 25 percent less than 4,000 deliveries.

In Use: **Have measure results been reported at this level previously?**
No.

Reliability & Validity: **Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?**
No.

Unintended consequences: **What are the potential unintended consequences of reporting at this level of aggregation?**
None anticipated.

Provider Level
Individual practitioner: Can compare individual health care professionals

Intended use: **Is measure intended to support meaningful comparisons at this level? (Yes/No)**
No.

Data Sources: **Are data sources available to support reporting at this level?**
No.

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**
Not specified for this purpose.

In Use: **Have measure results been reported at this level previously?**
No.

Reliability & Validity: **Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?**
No.

Unintended consequences: **What are the potential unintended consequences of reporting at this level of aggregation?**
Not applicable.

Provider Level
Hospital: Can compare hospitals

Intended use: **Is measure intended to support meaningful comparisons at this level? (Yes/No)**
No.

Data Sources: **Are data sources available to support reporting at this level?**
No.

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**
Not specified for this purpose.

In Use: **Have measure results been reported at this level previously?**
No.

Reliability & Validity: **Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?**
No.

Unintended consequences: **What are the potential unintended consequences of reporting at this level of aggregation?**
Not applicable.

Provider Level
Practice, group, or facility:** Can compare: (i) practice sites; (ii) medical or other professional groups; or (iii) integrated or other delivery networks

Intended use: **Is measure intended to support meaningful comparisons at this level? (Yes/No)**
No.

Data Sources: **Are data sources available to support reporting at this level?**
No.

Sample Size: **What is the typical sample size available for each unit at this level? What proportion of units at this level of aggregation can achieve an acceptable minimum sample size?**
Not specified for this purpose.

In Use: **Have measure results been reported at this level previously?**
No.

Reliability & Validity: **Is there published evidence about the reliability and validity of the measure when reported at this level of aggregation?**
No.

Unintended consequences: **What are the potential unintended consequences of reporting at this level of aggregation?**
Not applicable.

Section 10. Understandability

CHIPRA states that the core set should allow purchasers, families, and health care providers to understand the quality of care for children. Please describe the usefulness of this measure toward achieving this goal. Describe efforts to assess the understandability of this measure (e.g., focus group testing with stakeholders).

The focus of this CAPQuaM measure is women's exposure to teratogenic drugs before and throughout pregnancy, which is a marker for failure of preconception (inter-conception, where appropriate) care. At the individual level, the use of teratogens during pregnancy represents a failure of patient safety. While surely there are individual circumstances for which the anticipated benefits of use of a Class X medication is greater than the anticipated risk, by definition such circumstances are rare and should be considered to be extenuating. We consider variations in use of these medications at the population level to represent a marker for differences in the availability of safe and effective preconception high-risk obstetrical services.

We have not tested combining these measures into an index but can imagine some States or other entities wanting to do that. We will consider that in future work.

Understandability is at the heart of CAPQuaM's measure development process. Throughout development, CAPQuaM brought together diverse stakeholders – clinicians, scientists, payers, purchasers, consumer organizations, and others – to ensure their iterative engagement in advancing quality measures that are understandable, salient and actionable. CAPQuaM employed a 360 degree method, designed to involve key stakeholders in meaningful ways.

Our development process for this measure cultivated formal input from:

- Medical literature (both peer-reviewed and gray, including State Websites).
- Relevant clinicians.
- Organizational stakeholders (our consortium partners, as well as advisory board members).
- Multi-disciplinary, geographically diverse expert panel, including clinicians and academicians).
- CAPQuaM's scientific team.

Clinical criteria, including consideration of inclusion and exclusion criteria, were developed using a version of the RAND/UCLA modified Delphi Panels. CAPQuaM sought recommendations from major clinical societies and other stakeholders to identify academic and clinician expert panel participants with knowledge and expertise in a variety of areas and backgrounds, in both clinical and regional settings. The product of this process was participation by a broad group of experts in the development of clinically detailed scenarios leading to the measures.

CAPQuaM integrated perspectives from a national consortium, Steering Committee, and Senior Advisory Board at each step of the process, in addition to a continuing collaboration with AHRQ. Our team far exceeded the required minimums for expertise outside of the mainstream medical system, ensuring understandability at various levels and by a variety of audiences.

Alpha testing was performed to assess feasibility, mechanisms of data collection, and operational aspects of collecting and analyzing data for the measure.

Beta testing was performed by the New York State Office of Health Insurance Programs (Medicaid) in close collaboration with the CAPQuaM team.

The route to measure specification included development of relevant scenarios and issues for formal processing by our expert panel. Panel members participated in a two-round RAND/UCLA modified Delphi panel that culminated in a full 2-day in-person meeting that was moderated by a pediatrician and an obstetrician-gynecologist. The output from that panel meeting was summarized in the form of a boundary guideline that was then used to guide the measure specification and prioritization.

Section 11. Health Information Technology

Please respond to the following questions in terms of any health information technology (health IT) that has been or could be incorporated into the measure calculation.

11.A. Health IT Enhancement

Please describe how health IT may enhance the use of this measure.

As health information systems advance, perhaps the administrative data at the heart of this measure could migrate from billing and management systems to the electronic health record (EHR). We are not yet there.

11.B. Health IT Testing

Has the measure been tested as part of an electronic health record (EHR) or other health IT system?

No.

If so, in what health IT system was it tested and what were the results of testing?

Not applicable.

11.C. Health IT Workflow

Please describe how the information needed to calculate the measure may be captured as part of routine clinical or administrative workflow.

Other than perhaps the race/ethnicity data, the clinical data are a part of routine administrative data systems. The migration of diagnosis data from the EHR directly to administrative systems could conceivably improve the accuracy of the data in the future, although this is not clear.

11.D. Health IT Standards

Are the data elements in this measure supported explicitly by the Office of the National Coordinator for Health IT Standards and Certification criteria (see healthit.hhs.gov/portal/server.pt/community/healthit_hhs_gov__standards_ifr/1195)?

No.

If yes, please describe.

Not applicable.

11.E. Health IT Calculation

Please assess the likelihood that missing or ambiguous information will lead to calculation errors.

Not applicable.

11.F. Health IT Other Functions

If the measure is implemented in an EHR or other health IT system, how might implementation of other health IT functions (e.g., computerized decision support systems in an EHR) enhance performance characteristics on the measure?

Not applicable.

Section 12. Limitations of the Measure

Describe any limitations of the measure related to the attributes included in this CPCF (i.e., availability of measure specifications, importance of the measure, evidence for the focus of the measure, scientific soundness of the measure, identification of disparities, feasibility, levels of aggregation, understandability, health information technology).

Our definition of high-risk obstetrical care is based on a careful, evidence-driven consensus process that was highly engaged and guided by an extraordinary multidisciplinary panel of national experts. The CAPQuaM team carefully operationalized their conclusions and maintained dialogue as we did so. Still there were infinite combinations of qualifying criteria, and we had to specify one. We are confident that the specifications are strong, the conditions meaningful, and the population is at increased risk. But these were designed from the outset and explicitly discussed at the expert meeting to be population-based measures. They are intended for the measurement of performance across populations, not for the assessment of the quality of an individual's care. The inevitable noise in the measures was designed to be dwarfed by the signal when applied to large numbers of pregnant women but not for any given individual.

This measure requires prescription fill data specified in State Medicaid records, health plans, and other administrative data sources. Our colleagues at the New York State Department of Health and other members of our Steering Committee have confirmed that this is a feasible and valid way to assess teratogenic drug exposure. However, because of limitations in the current data system, we can only assess prescription fill at this time rather than the actual use of such medications. Finally, our inclusion of only Class X medications enhances the sensitivity of this meaning for identifying avoidable pharmacological risk during pregnancy at the expense of sensitivity.

Section 13. Summary Statement

Provide a summary rationale for why the measure should be selected for use, taking into account a balance among desirable attributes and limitations of the measure. Highlight specific advantages that this measure has over alternative measures on the same topic that were considered by the measure developer or specific advantages that this measure has over existing measures. If there is any information about this measure that is important for the review process but has not been addressed above, include it here.

This innovative measure addresses a complex and critical idea: How available is the preconception component of high-risk obstetrical services by estimating the use of teratogenic medications before and during pregnancy. Specifically, how often are women exposed to potentially harmful teratogens during their pregnancy because they lack availability of preconception or inter-conception care. We set forth specifications to identify pregnancies that constitute high risk and those that are exposed to teratogens.

These measures respond to the assignment to CAPQuaM, an AHRQ-CMS CHIPRA Center of Excellence in the Pediatric Quality Measurement Program. We have used a rigorous and systematic process that was highly engaged with clinicians, stakeholders, and experts to develop these measures. We began with the evidence base and the literature.

Four million births occur annually, and our data demonstrate that in any given State, between one- and two-thirds of pregnancies are high risk. Exposure to teratogens during pregnancy is associated with significant adverse outcomes and occurs with some frequency. The rapidly rising rate of teratogenic drug use and associated complications highlight the need for increased availability of preconception (inter-conception) HROB care.

These are important measures regarding quality and patient safety. Given the number and severity of fetal effects that can occur with many different medications, discussing medication use with women of reproductive age and child-bearing ability is critical, as is avoiding the concurrence of pregnancy with teratogenic medications whenever possible. However, data suggest that counseling occurs in only one-quarter to one-half of cases. There is a growing need to address preconception care as a means of reducing risk among women and their offspring. This measure addresses one component of preconception care by estimating the use of teratogenic medications before and during pregnancy. It is both a safety measure for these women and a population marker for preconception care more generally.

Our validation tests showed that rates of teratogen use vary by race and geography. We found the sub-measures to be complementary and not duplicative. The sub-measures were sensitive to differences in race, and urbanicity. We found they could be implemented in New York State Medicaid data. The measures performed well.

References

Adam M, Polifka JE, Friedman JM. Evolving knowledge of the teratogenicity of medications in human pregnancy. Am J Med Genet C Semin Med Genet 2011; 157(3):175-82.

Akers A, Gold MA, Borrero S, et al. Providers' perspectives on challenges to contraceptive counseling in primary care settings. J Womens Health 2010; 19(6):1163-70.

Bennett KJ, Olatosi B, Probst JC. Health disparities: A rural-urban chartbook. Columbia, SC: South Carolina Rural Health Research Center; 2008.

Broussard CS, Frey MT, Hernandez-Diaz S, et al., Developing a systematic approach to safer medication use during pregnancy: Summary of a Centers for Disease Control and Prevention–convened meeting. Am J Obstet Gynecol 2014; 211(3):208-14.

Burkey B, Holmes AP. Evaluating medication use in pregnancy and lactation: What every pharmacist should know. J Pediatr Pharmacol Ther 2013; 18(3):247-58.

Centers for Disease Control and Prevention (CDC). Facts about birth defects, 2014; fact sheet. Atlanta, GA: CDC; updated 2017. Available at https://www.cdc.gov/ncbddd/birthdefects/facts.html. Accessed May 30, 2018.

Centers for Disease Control and Prevention (CDC). Treating for two: A national strategy for medication use in Pregnancy. Atlanta, GA: CDC; 2014. Available at https://www.cdc.gov/ncbddd/birthdefects/documents/ncbddd_birth-defects_medicationuseonepager_cdcrole.pdf. Accessed May 30, 2018.

Children's Health Insurance Program Reauthorization Act of 2009 (CHIPRA). Public Law 111-3, 123 Stat. 8 (2009). Available at https://www.gpo.gov/fdsys/pkg/PLAW-111publ3/html/PLAW-111publ3.htm. Accessed June 4, 2018.

Cooper W, Hickson GB, Ray WA. Prescriptions for contraindicated category X drugs in pregnancy among women enrolled in TennCare. Paediatr Perinat Epidemiol 2004; 18(2):106-11.

Daw J, Hanley GE, Greyson DL, et al. Prescription drug use during pregnancy in developed countries: A systematic review. Pharmacoepidemiol Drug Saf 2011; 20(9): 895-902.

Diamond CC, Rask KJ, Kohler SA. Use of paper medical records versus administrative data for measuring and improving health care quality: Are we still searching for a gold standard? Dis Manag 2004; 4(3):121-9.

Eisenberg D, Stika C, Desai A, et al. Providing contraception for women taking potentially teratogenic medications: A survey of internal medicine physicians' knowledge, attitudes and barriers. J Gen Intern Med 2010; 25(4):291-7.

Food and Drug Administration. Pregnancy and lactation labeling (Drugs) final rule. December 2014. Available at https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/DrugInteractionsLabeling/ucm093307.htm. Accessed May 30, 2018.

Force RW, Keppel GA, Guirguis-Blake J, et al. contraceptive methods and informed consent among women receiving medications with potential for adverse fetal effects: A

Washington, Wyoming, Alaska, Montana, Idaho (WWAMI) region study. J Am Board Fam Med 2012; 25(5):661-8.

Gawron L, Hammond C, Keefer L. Documentation of reproductive health counseling and contraception in women with inflammatory bowel diseases. Patient Educ Counsel 2014; 94(1):134-7.

Guttmacher Institute. Unintended pregnancy in the United States, 2013. Guttmacher Institute. Washington, DC: Guttmacher Institute; updated 2016. Available at https://www.guttmacher.org/fact-sheet/unintended-pregnancy-united-states. Accessed May 30, 2018.

Hart G. Frontier/Remote, Island, and Rural Literature Review. Grand Forks, ND: University of North Dakota; 2012: Available at https://ruralhealth.und.edu/pdf/frontierreview.pdf.Accessed August 7, 2018.

Huybrechts K, Palmsten K, Hernandez-Diaz S, et al. Antidepressant use in pregnancy and the risk of cardiac defects. New Engl J Med 2014; 370(25):2397-407.

Institute of Medicine. Crossing the Quality Chasm: The IOM health care quality initiative. Washington, DC: National Academies Press; 2001.

Jentink J, Dolk H, Loane MA, et al., Intrauterine exposure to carbamazepine and specific congenital malformations: Systematic review and case-control study. BMJ 2010; 341:c6581.

Jentink J, Loane MA, Dolk H, et al. Valproic acid monotherapy in pregnancy and major congenital malformations. New Engl J Med 2010; 362(23):2185-93.

Kuhlthau, K. Measures of availability of health care services for children. Acad Pediatr 2011; 11(3 Suppl):S42-8.

Kuklina E, Callaghan W. Chronic heart disease and severe obstetric morbidity among hospitalisations for pregnancy in the USA: 1995-2006. BJOG 2011; 118(3):345-52.

Langan J, Perry A, Oto M. Teratogenic risk and contraceptive counselling in psychiatric practice: Analysis of anticonvulsant therapy. BMC Psychiatry 2013. 13(1):234.

Lo W, Friedman JM. Teratogenicity of recently introduced medications in human pregnancy. Obstet Gynecol 2002; 100(3):465-73.

Mangione-Smith R, DeCristofaro AH, Setodji CM, et al. The quality of ambulatory care delivered to children in the United States. New Engl J Med 2007; 357(15):1515-23.

Mitchell A, Gilboa SM, Werler MM, et al. Medication use during pregnancy, with particular focus on prescription drugs: 1976-2008. Am J Obstet Gynecol 2011; 205(1):51.e1-8.

Peters SL, Lind JN, Humphrey JR, et al. Safe lists for medications in pregnancy: Inadequate evidence base and inconsistent guidance from Web-based information, 2011.

Pharmacoepidemiol Drug Saf 2013; 22(3):324-8.

Rubio D, Berg-Weger M, Tebb SS, et al. Objectifying content validity: Conducting a content validity study in social work research. Soc Work Res 2003; 27(2):94-104.

Santucci A, Gold MA, Akers AY, et al. Women's perspectives on counseling about risks for medication-induced birth defects. Birth Defects Res A Clin Mol Teratol 2010; 88(1):64-9.

Schwarz EB, Longo LS, Zhao X, et al. Provision of potentially teratogenic medications to female veterans of childbearing age. Med Care 2010; 48(9):834-42.

Schwarz E., Maselli J, Norton M, et al. Prescription of teratogenic medications in United States ambulatory practices. Am J Med 2005; 118(11):1240-9.

Schwarz E, Parisi SM, Handler SM, et al. Clinical decision support to promote safe prescribing to women of reproductive age: A cluster-randomized trial. J Gen Intern Med 2012; 27(7):831-8.

Schwarz E, Parisi SM, Handler SM, et al. Counseling about medication-induced birth defects with clinical decision support in primary care. J Womens Health 2013; 22(10): 817-24.

Schwarz E, Postlethwaite DA, Hung Y, et al. Documentation of contraception and pregnancy when prescribing potentially teratogenic medications for reproductive-age women. Ann Intern Med 2007; 147(6):370-6.

Schwarz E, Postlethwaite DA, Hung Y, et al. Documentation of contraception and pregnancy when prescribing potentially teratogenic medications for reproductive-age women. Ann Intern Med 2007; 147(6):370-6.Schwarz E, Santucci A, Borrero S, et al. Perspectives of primary care clinicians on teratogenic risk counseling. Birth Defects Res A Clin Mol Teratol 2009; 85(10):858-63.

Steinkellner A, Chen W, Denison SE. Adherence to oral contraception in women on category X medications. Am J Med 2010; 123(10):929-34.e1.

Taruscio D, Arriola L, Baldi F, et al. European recommendations for primary prevention of congenital anomalies: A joined effort of EUROCAT and EUROPLAN projects to facilitate inclusion of this topic in the National Rare Disease Plans. Public Health Genomics 2014; 17(2):115-23.

Virnig B, McBean M. Administrative data for public health surveillance and planning. Annu Rev Public Health 2001; 22:213-30.

Section 14: Identifying Information for the Measure Submitter

First Name:	Lawrence
Last Name:	Kleinman
Title:	Director, Mount Sinai CAPQuaM
Organization:	Icahn School of Medicine at Mount Sinai
Mailing Address:	One Gustave L. Levy Place, Box 1077
City:	New York
State:	NY
Postal Code:	10029
Telephone:	212-659-9556
Email:	lawrence.kleinman@mssm.edu

The CHIPRA Pediatric Quality Measures Program (PQMP) Candidate Measure Submission Form (CPCF) was approved by the Office of Management and Budget (OMB) in accordance with the Paperwork Reduction Act.

The OMB Control Number is 0935-0205 and the Expiration Date is December 31, 2015.

Public Disclosure Requirements

Each submission must include a written statement agreeing that, should U.S. Department of Health and Human Services accept the measure for the 2014 and/or 2015 Improved Core Measure Sets, full measure specifications for the accepted measure will be subject to public disclosure (e.g., on the Agency for Healthcare Research and Quality [AHRQ] and/or Centers for Medicare & Medicaid Services [CMS] websites), except that potential measure users will not be permitted to use the measure for commercial use. In addition, AHRQ expects that measures and full measure specifications will be made reasonably available to all interested parties. "Full measure specifications" is defined as all information that any potential measure implementer will need to use and analyze the measure, including use and analysis within an electronic health record or other health information technology. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure. This statement must be signed by an individual authorized to act for any holder of copyright on each submitted measure or instrument. The authority of the signatory to provide such authorization should be described in the letter.

AHRQ Publication No. 17(18)-P008-2-EF
August 2018